Regression Coursework

Hoe Keng Yew

15/10/2020

Summary

This report analyzed the relationship between transmission types and miles per gallon (MPG). The aim of this investigation is to determine which transmission type is more fuel economical (higher MPG value). After the analysis was done, manual transmission cars generally have a higher MPG value however the difference in MPG between automatic and manual transmission is only 1.81. Other factors such as number of cylinders, horsepower and weight of the car have a more significant effect on the MPG value.

Loading the data

```
# loading the R Built-in dataset
library(ggplot2)
data(mtcars)
mtcars$mpg <- as.numeric(mtcars$mpg)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- ifelse(as.integer(mtcars$am == 1), "Manual", "Auto")

head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs     am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0 Manual    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0 Manual    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1 Manual    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1   Auto    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0   Auto    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1   Auto    3    1
```

Exploratory Analysis

Regression model

```
fullModel <- lm(mpg~., mtcars)
optimalModel <- step(fullModel, direction = "backward")
```

This procedure determines that the best model includes the cyl6, cyl8, hp, wt, and amManual variables (overall p-value<0.001). The adjusted R-squared indicates that about 84% of the variance is explained by the final model. Moreover, the output of this model suggests that mpg decreases with respect to cylinders (-3.03 and -2.16 for cyl6 and cyl8, respectively), horsepower (-0.03), and weight (for every 1,000lb, by -2.5). On the other hand, mpg increases with respect to having a manual transmission (by 1.8). Residual plots (see appendix) suggest that some transformation may be necessary to achieve linearity.

Statistical Inference

```
t.test(mpg~am,mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
##   mean in group Auto mean in group Manual
##             17.14737             24.39231
```

The confidence intervals excludes 0 and the p-value is greater than the threshold of 0.05. The null hypothesis can be rejected.

Conclusion

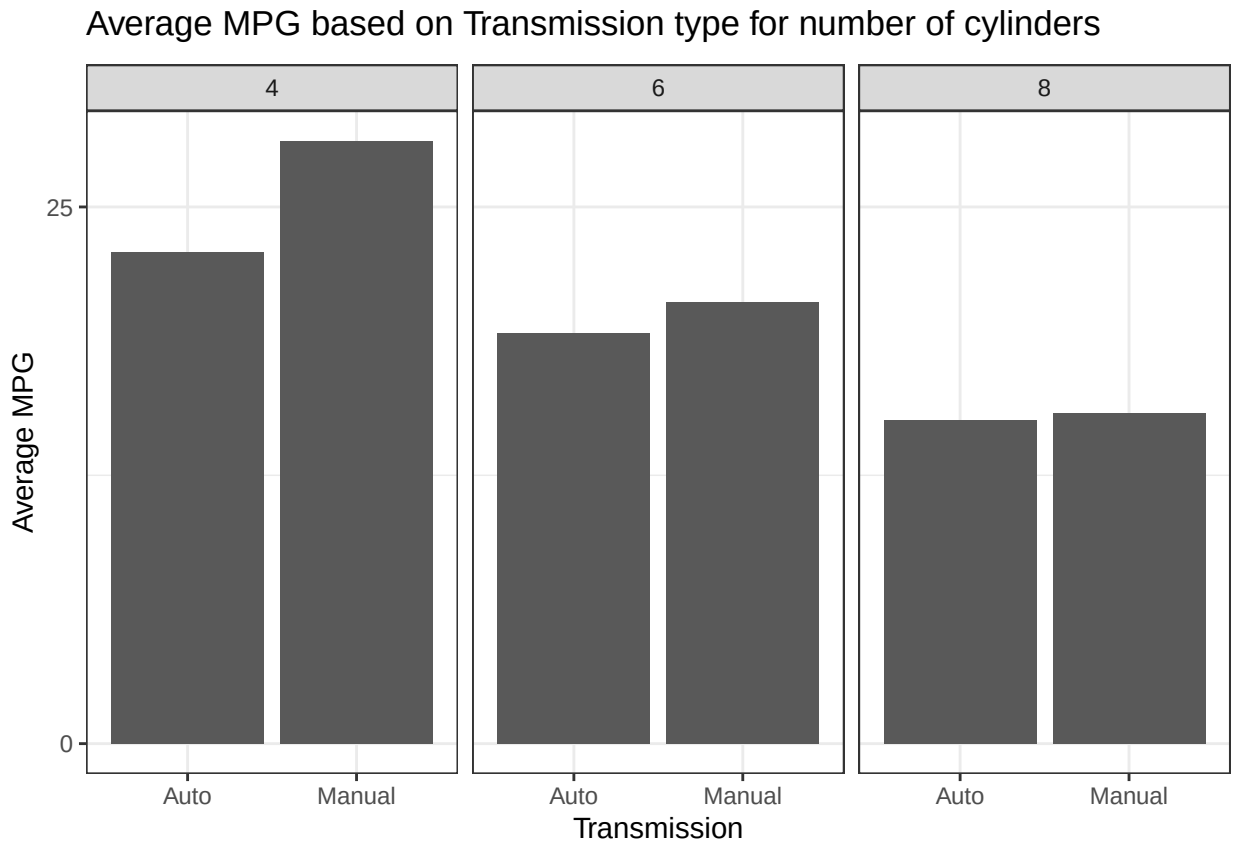
According to the data analysis, cars with manual transmission have better fuel efficiency as they have a higher MPG value. The hypothesis testing done for the condition where MPG is the same for automatic transmission and manual transmission are rejected. Cars with a lower number of cylinders have a higher MPG value compared to cars with higher number of cylinders. The difference between MPG values for Automatic and Manual transmission cars also decreases as umber of cylinders increase.

Appendix

Graphs and Summaries of tables

```
# Finding the average mpg for auto and manual transmission cars based on cylinders
transmissionMPG <- aggregate(mpg ~ am + cyl, mtcars, FUN = mean )
#head(transmissionMPG)

# Plotting the histogram
p <- ggplot(transmissionMPG, aes(x = am,  y = mpg)) + geom_bar(stat = "identity") +facet_grid(scales="fi
  guides(fill=FALSE) + theme_bw() +
  scale_fill_brewer(palette="Blues")+labs(x="Transmission", y=expression("Average MPG")) +
  labs(title=expression("Average MPG based on Transmission type for number of cylinders")) + scale_y_co
p
```

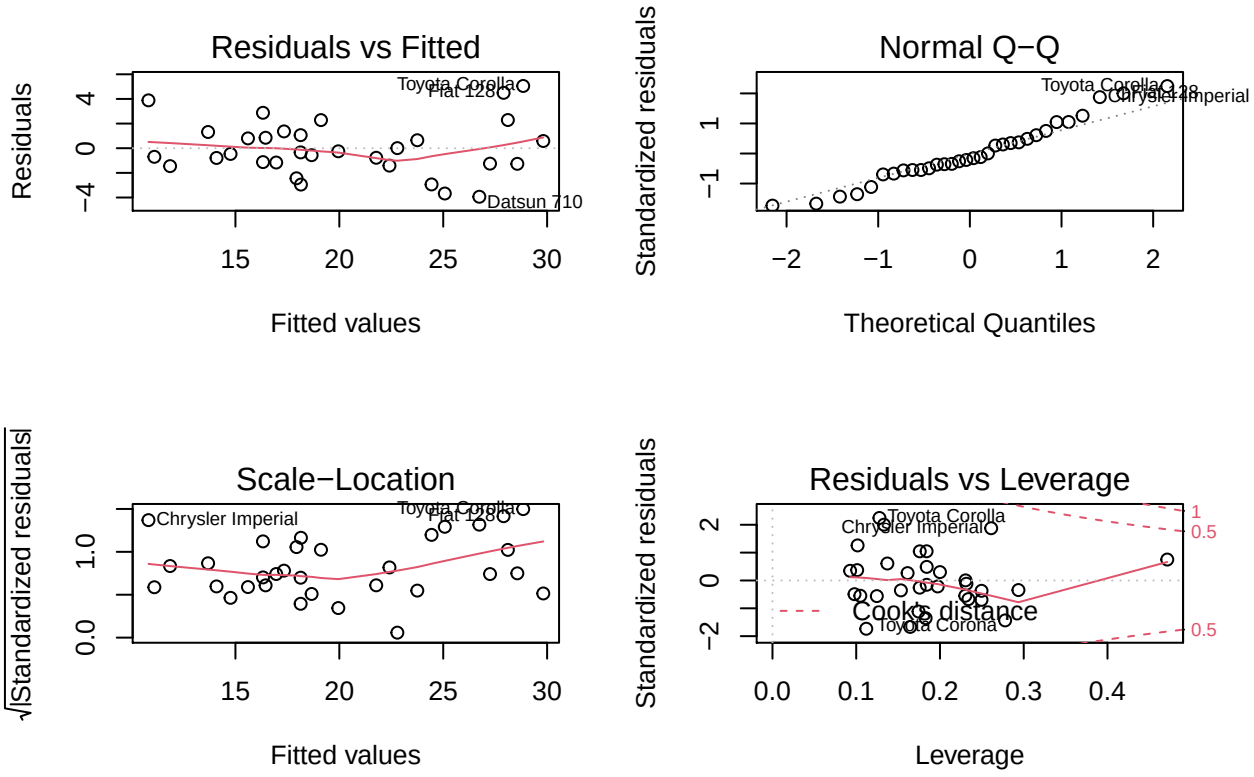



```
# Summary of the best model
summary(optimalModel)
```

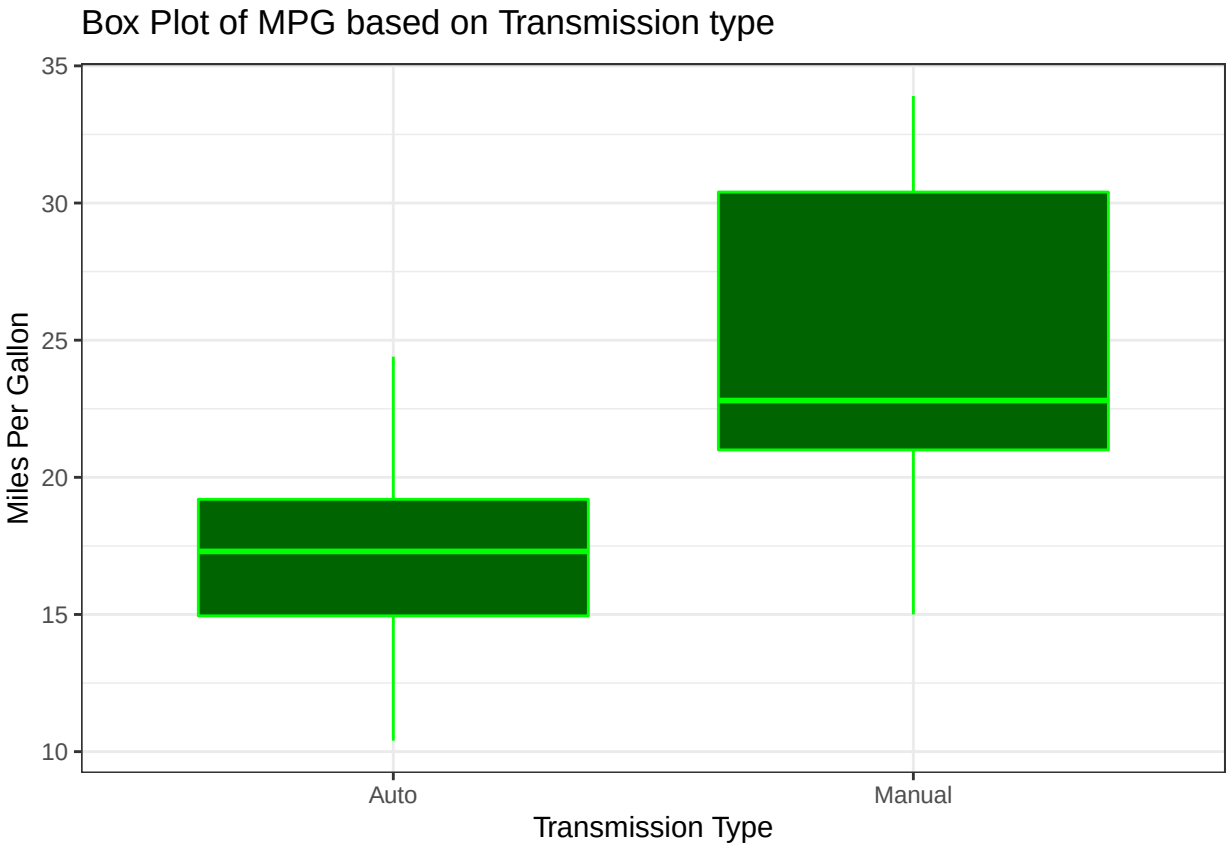
```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

```
par(mfrow=c(2,2))
plot(optimalModel)
```




```
# Box Plot of the MPG based on transmission type
b<-ggplot(data = mtcars,aes(x=am,y=mpg)) + geom_boxplot(col = "green", fill = "darkgreen") + theme_bw()
b
```


Box Plot of MPG based on Transmission type
35
30
25
20
15
10
Miles Per Gallon
Auto
Manual
Transmission Type